European Journal of Operational Research 270 (2018) 352–361

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

A non-compensatory composite indicator approach to assessing low-carbon performance

L. P. Zhang[a], P. Zhou[a,b,*]

[a] College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing, China
[b] School of Economics and Management, China University of Petroleum, No. 66 Changjiang West Road, Qingdao, China

ARTICLE INFO

Article history:
Received 26 July 2017
Accepted 27 February 2018
Available online 7 March 2018

Keywords:
OR in environment and climate change
Composite indicator
Low-carbon performance
Aggregation
Non-compensatory

ABSTRACT

Low-carbon development has been widely regarded as a key strategy for tackling the challenges posed by climate change. Measuring low-carbon performance can provide policy makers valuable information for monitoring the progress of low-carbon development in an economy such as a city. Composite indicator, owing to its transparency and ease of communication to the public, has been touted as a useful analytical tool for measuring low-carbon performance. The construction of composite indicators often takes the compensability assumption which allows the full substitutability between underlying indicators. In this paper, we argue that the compensability assumption needs to be restricted in assessing low-carbon performance. A non-compensatory approach based on the outranking relation is used to construct composite low-carbon performance indicator. A more efficient heuristic procedure is proposed to handle the computational complexity in deriving the final comprehensive rankings. The approach has been applied to assess the city-level low-carbon performance in China. A sensitivity analysis is conducted to investigate the impacts of various parameters on the modeling results.

1. Introduction

Decarbonization has been regarded as a key strategy for tackling the environmental and economic challenges posed by climate change in many countries (IPCC, 2015). International communities have devoted to searching for low-carbon development path in order to achieve the joint goals of economic growth and CO_2 emission reduction. In this context, the measurement of low-carbon performance is timely and valuable as it can provide policy makers useful information for comparing and monitoring the progress of different entities towards low-carbon development. Several earlier studies have contributed to develop methodological frameworks for evaluating low-carbon performance. An example is Price et al. (2013) who used a set of indicators to assess the low-carbon performance of different provinces in China from a sectoral perspective. Other studies, e.g. Yu (2014), Zhou, He, Williams, and Fridley (2015) and Tan et al. (2017), took a more integrative perspective in evaluating low-carbon performance. Their methods often involve multiple steps such as indicator selection, data normalization, weighting and aggregation, which are theoretically consistent with the methodological framework for constructing composite indicators as described by JRC (2008).

Composite indicator (CI) is an analytical tool that aims to summarize complex, multi-dimensional low-carbon performance into a performance value for ease of interpretation and communication. Broadly speaking, we can categorize the steps for constructing CIs into two main stages. The first stage involves the determination of a suitable indicator framework that can capture the main characteristics of low-carbon development. The second stage involves the assignment of appropriate weights to underlying indicators and their aggregation into a composite low-carbon performance indicator. Despite the importance of determining indicator framework and assigning weights, in this paper we shall only focus on data aggregation for which multi-criteria techniques play a vital role.[1] For example, the simple additive weighting (SAW) method has been widely used in practice owing to its merits such as transparency and ease of communication. Despite its popularity, the SAW method is unlikely to generate a meaningful CI in many cases as discussed by Ebert and Welsch (2004) and Böhringer and Jochem (2007). In addition, Zhou, Ang, and Poh (2006) showed that

* Corresponding author at: College of Economics and Management, Nanjing University of Aeronautics and Astronautics, 29 Jiangjun Avenue, Nanjing, China.
E-mail address: cemzp@nuaa.edu.cn (P. Zhou).

[1] Pollesch and Dale (2015) provided a review of alternative aggregation methods for constructing composite indicators. More recently, Diaz-Balteiro, González-Pachón, and Romero (2017) carried out a critical and comprehensive review on the applications of multi-criteria methods to measuring sustainability performance.

https://doi.org/10.1016/j.ejor.2018.02.058
0377-2217/

the SAW method may lead to more information loss compared to several other aggregation methods. Diaz-Balteiro, González-Pachón, and Romero (2017) pointed out that the assumption of preference independence among individual indicators or criteria underlying the SAW method needs to be considered in application. Accordingly, other multi-criteria techniques, ranging from the weighted product (WP) method to mathematical programming models, have been advocated for use. Examples of such studies can be found in Despotis (2005), Tofallis (2013, 2014), Van Puyenbroeck and Rogge (2017), Zanella, Camanho, and Dias (2015), Zhou, Ang, and Poh (2007, 2010, 2012), Rogge (2018) and Verbunt and Rogge (2018). Zhou, Delmas, and Kohli (2017) extended the concept of meaningfulness initiated by Ebert and Welsch (2004) and highlighted the importance of preserving the relative performance gaps in CI construction. Fusco, Vidoli, and Sahoo (2017) proposed an approach to constructing CI when exogenous external factors are under consideration.

It should be pointed out that the linear aggregation methods such as SAW implicitly take the full compensatory assumption (Diaz-Balteiro et al., 2017; Fusco, 2015). According to Munda and Nardo (2009), full compensability is often undesirable from a normative perspective. The non-linear aggregation methods have thus been used in some cases, e.g. Van Puyenbroeck and Rogge (2017) and Verbunt and Rogge (2018). A well-known CI is the Human Development Index that is currently constructed by the weighted production (WP) method as a replacement of the SAW method (Tofallis, 2013). When the WP method is used, the compensability between indicators is variable rather than fixed (Van Puyenbroeck & Rogge, 2017). In spite of the theoretical strengths, the WP method still shows certain compensability, which indicates that the inferiority in one indicator can still be offset by a superiority in another indicator. In evaluating low-carbon performance, it may not be appropriate to allow for the substitutability between environmental and economic indicators (Díaz-Balteiro & Romero, 2004; Mori & Christodoulou, 2012). In addition, Munda and Nardo (2005, 2009) showed that the interpretation of the weights for underlying indicators in the compensatory aggregation methods is theoretically inconsistent with their original meaning of importance coefficients, especially when the preferential independence between these indicators cannot be confirmed.[2]

In this paper, we argue that the compensatory assumption needs to be restricted in evaluating low-carbon performance and propose a non-compensatory approach for the purpose. The non-compensatory approach has the flexibility of integrating both quantitative and qualitative indicators and allows for the evaluation results to be reported in their natural measurement units. Through the use of non-compensatory approach, the composite low-carbon performance indicators derived may be more informative and intuitive to policy makers and public. Theoretically, there are different non-compensatory methods, e.g. dominance analysis, satisficing methods, sequential elimination methods, attitude oriented methods and outranking relation analysis (Yoon & Hwang, 1995). Some refined goal programming models can also alleviate the compensation among indicators, e.g. Blancas, Caballero, González, Lozano-Oyola, and Pérez (2010). Following Munda and Nardo (2005, 2009), in this paper we use the outranking relation to develop a non-compensatory composite indicator for evaluating low-carbon performance. An outranking relation is defined as a binary relation S based on pairwise comparisons of entities to be evaluated on different criteria, where xSy indicates that entity x is at least as good as y. More details on the description of outranking relation and its application studies can be found in Roy (1991) and Yoon and Hwang (1995) and the survey study by Govindan and Jepsen (2016).

The methodological contribution of this paper mainly lies in two aspects. First, we integrate preference thresholds into the non-compensatory CI framework that helps to increase the discriminating power in constructing outranking matrix. Meanwhile, it prevents that local indifference results in global indifference and thus enhances the robustness of the evaluation results. Second, we introduce a more efficient heuristic procedure for deriving the final comprehensive rankings of entities, which helps to address the computational complexity occurred in the non-compensatory approach as described in Munda and Nardo (2009). In application, we apply the proposed approach to evaluating the low-carbon performance of forty cities in China and present the results obtained. A sensitivity analysis is also conducted to investigate the robustness of the evaluation results with respect to various parameters.

The remainder of this paper is organized as follows. In Section 2, we present the non-compensatory approach. In Section 3, we present the results of the empirical study on evaluating the low-carbon performance of forty cities in China. Section 4 concludes this study.

2. Methodology

Suppose that $A = [a,\ b,\ \dots]_n$ represents a set of entities whose low-carbon performance is to be evaluated based on M indicators g_j $(j = 1,\ \dots,\ M)$ with different measurement units. $g_j(a)$ denotes the performance of a with respect to the jth indicator. We also assume that all the indicators are of the benefit type. If $gj(a) > gj(b)$, entity a performs better than entity b on the jth indicator. If $gj(a) = gj(b)$, the performance of a and b would be indifferent with respect to indicator j. Further assume that an importance coefficient wj (i.e. weight) is assigned to indicator g_j. The purpose is to construct a CI for evaluating the comprehensive low-carbon performance of each entity. The common aggregation approaches such as SAW and WP aim to elaborate a value function $V(a) = V[gj(a)]$, based on which the entities' performance can be directly evaluated and compared. For example, the performance of a and b would be indifferent if $V(a) = V(b)$. If $V(a) > V(b)$, the performance of a would be superior to b.

The study by Munda and Nardo (2009) provides an intuitive non-compensatory multi-criteria framework for constructing CI, which involves two steps in order to establish a comprehensive preference for each entity. Firstly, an outranking matrix is built by pair-wise comparisons of all the entities in term of indicator family based on the outranking relation concept. The element *eab* in the outranking matrix indicates that the performance of entity a, with respect to the indicator family, is at least as good as that of entity b. The second step is to rank the entities based on a complete pre-order. The framework provides a basis for constructing a non-compensatory CI for evaluating low-carbon performance. Despite its theoretical strengths, the framework might face the Luce's paradox of "the cup of coffee" (Luce, 1956) which is illustrated through the example of air quality index as follows. In reality, few people can distinguish the difference between a city with air quality index value $g(a)$ and another city with the value $g(a) + 0.01$. We may remain indifferent if the value of air quality index is added by 0.01 at a time. Consequently, when the indifference relation is transitive, we may draw the conclusion that a city with poor air quality would be indifferent from a city with good air quality, which obviously makes no sense. In practice, the error arising from data collection and compilation can also make the preference/indifference framework be problematic.

To alleviate the above paradox and improve the robustness of the multi-criteria framework with respect to data uncertainty, we here introduce the concept of threshold based on which

[2] It should be pointed out that some semi-compensatory CI frameworks can keep the consistency by using budget share method to assign indicator weights. See, for example, Van Puyenbroeck and Rogge (2017).

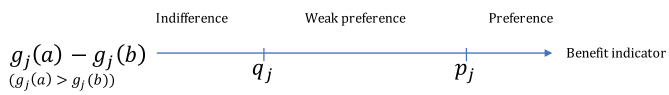

Fig. 1. The preference relation with consideration of thresholds.

$gj(a) > gj(b)$ will no longer strictly support the statement of a performing better than b. Alternatively, we need to examine whether the positive difference $gj(a) - gj(b)$ is sufficiently higher than a positive preference threshold, i.e. pj. Formally speaking, the preference relation between a and b with respect to indicator g_j needs to satisfy the following condition: $g_j(a) > g_j(b)$ and $|g_j(a) - g_j(b)| > p_j$. Analogously, to verify the statement of a performing indifferent from b, $|gj(a) - gj(b)|$ should be sufficiently less than or equal to a positive indifference threshold, i.e. $|gj\ (a) - gj\ (b)| \leq qj$. When the difference falls into the range between qj and, their relationship will be described as weak preference. Fig. 1 provides a graphical illustration of the relationship.

The concept of thresholds introduced here can prevent that local indifference results in global indifference due to transitivity. It should be pointed out that the threshold concept is not new and has rooted from the non-compensatory multiple attribute decision analysis (Yoon & Hwang, 1995). For example, Roy (1991) elaborated this concept as the foundation of the ELECTRE family. More recently, Rowley, Peters, Lundie, and Moore (2012) suggested to incorporate thresholds in sustainability evaluation and comparison. Truong, Adamowicz, and Boxall (2015) applied a similar concept in environmental valuation. Attardi, Cerreta, Sannicandro, and Torre (2018) used the indicator thresholds when evaluating land-use policy efficiency.

Based on thresholds, the indicator set can be divided into two exclusive subsets. One is the subset verifying the statement of a's performance being at least as good as that of b in terms of certain indicators, including the indicators representing indifference and preference relations (relation S). The other subset is composed of the remaining indicators representing weak preference (relation Q). The former subset satisfies the requirement of the outranking matrix in the non-compensatory multi-criteria framework proposed by Munda and Nardo (2009). However, the indicators in the latter subset also have the possibility to contribute their share to the outranking matrix since the weak preference reflects the hesitation between the indifference relation and the preference relation (Roy, Figueira, & Almeida-Dias, 2014). Hence, the elements of the outranking matrix are composed of both strong outranking relation (S) and weak outranking relation (part of Q). Based on the reasoning, we modify the first step of the non-compensatory multi-criteria framework by Munda and Nardo (2009) by including two more components, i.e.

$$e_{ab} = \sum_{j=1}^{M} w_j^P + \phi w_j^I + \psi w_j^Q \tag{1}$$

where w_j^P, w_j^I, and w_j^Q are the importance coefficient w of each indicator when the corresponding pair-wise indicator comparison shows the preference, indifference and weak preference, respectively. The coefficients ϕ and ψ represent the contributions of those indicators presenting indifference and weak preference relations supporting that a performing at least as good as b. Here we temporarily assume that $\phi = \psi = 0.5$ for ease of elaboration.

The numerical example given by Munda and Nardo (2009) is used to illustrate how to construct the outranking matrix with thresholds (see Table 1 for the dataset). The weights and thresholds are arbitrarily determined and shown in the last three rows of Table 1. Comparing A and B, for example, without consideration of thresholds, A performs better on three indicators (i.e. solid waste, income disparity and crime rate), and worse on the other two indicators (i.e. GDP and unemployment rate). Hence, the element e_{AB} of outranking matrix will be

$$\begin{aligned} e_{AB} &= w_{solide\ waste} + w_{income\ disparity} + w_{crime\ rate} \\ &= 0.333 + 0.165 + 0.165 \\ &= 0.666 \end{aligned}$$

With the thresholds, A and B becomes weak preference relation in terms of crime rate and indifferent in terms of solid waste and income disparity. Thus, the value of eAB becomes

$$\begin{aligned} e_{AB} &= \frac{1}{2}(w_{solid\ waste} + w_{income\ disparity}) + \frac{1}{2}w_{crime\ rate} \\ &= \frac{1}{2}(0.333 + 0.165) + \frac{1}{2}(0.165) \\ &= 0.333 \end{aligned}$$

Other elements in the outranking matrix can be calculated in a similar way.

Table 2 shows the outranking matrix with and without consideration of thresholds. It can be seen from Table 2 that the relationship between two entities is heavily influenced by the indicator thresholds, even though the small values of thresholds are used.

Next, we shall introduce the ranking procedure with the modified outranking matrix, which theoretically follows a maximum likelihood ranking principle. Technically, we should firstly determine all the possible permutations of the entities and then calculate the score for every permutation. The permutation with the highest score would be the final ranking. In the numerical example, all the possible permutations of the three entities as well as the evaluation results are shown in Table 3. It can be observed that the permutation $C\ B\ A$ achieves the highest score 1.830, which implies that C performs better than B that is followed by A. However, based on the original non-compensatory CI framework, the final ranking would be $C\ A\ B$. The controversy is caused by the comparison between A and B. Without consideration of the thresholds, it seems to be reasonable to rank A over B since A performs better on the last three indicators as shown in Table 1. Nevertheless, we may find that the performance difference between A and B with respect to the three indicators is very small, while the performance of B on the first two indicators is much better than that of A. With thresholds, the statement of B performing better than A seems to be more reasonable since the outranking relation between B and A becomes indifferent on solid waste and income disparity and weak preference on crime rate, but remains preference on the other two

Table 1
The performance matrix of the numerical example.

	GDP	Unemployment rate	Solid waste	Income disparity	Crime rate
Entity A	22000	0.17	0.40	10.50	40
Entity B	45000	0.09	0.45	11.0	45
Entity C	20000	0.08	0.35	5.30	80
w_j	0.165	0.165	0.333	0.165	0.165
q_j	1000	0.02	0.05	1	4
p_j	3000	0.04	0.10	2	10

Table 2
The outranking matrix.

(a) Without thresholds						(b) With thresholds					
	A	B	C	Strength score	Rank		A	B	C	Strength score	Rank
A	–	.666	.333	.999	2	A	–	.333	.417	.750	3
B	.333	–	.333	.666	3	B	.666	–	.417	1.083	2
C	.666	.666	–	1.332	1	C	.582	.582	–	1.164	1

Note: The columns of strength score and rank are generated by the heuristic ranking procedure described below.

Table 3
Evaluation results by the non-compensatory approach with and without thresholds.

Permutation	Calculation	Result
ABC	.333(AB) + 0.417(AC) + 0.417(BC)	1.167 (1.333)
ACB	.417(AC) + 0.333(AB) + 0.582(CB)	1.332 (1.666)
BAC	.666(BA) + 0.417(BC) + 0.417(AC)	1.500 (1.000)
BCA	.417(BC) + 0.666(BA) + 0.582(CA)	1.665 (1.333)
CAB	.582(CA) + 0.582(CB) + 0.333(AB)	1.497 (2.000)
CBA	.582(CB) + 0.582(CA) + 0.666(BA)	1.830 (1.666)

Note: The values given by the non-compensatory composite indicator approach without thresholds are shown in the parentheses.

indicators. The conclusion provides support to integrate thresholds into the non-compensatory CI framework.

In virtue of its non-compensatory feature, the above framework can integrate the distinct indicators in their natural units, which makes the results be more informative and intuitive. Furthermore, it is also flexible to aggregate quantitative as well as qualitative indicators in a meaningful way. Despite the merits, a challenge behind the use of the non-compensatory approach in practice is its large computational burden since the final ranking is based on the calculation for every possible permutation (Attardi et al., 2018). If n becomes rather large, it is almost infeasible to achieve a comprehensive ranking. Hence, it is necessary to simplify the ranking procedure.

We turn back to the outranking matrix in which the row vectors actually have the meaning of strength degree. For instance, eAB in Table 2 represents the degree of A performing at least as good as B. eAC would have a similar meaning. As a result, the sum of the row elements could be interpreted as the overall strength of an individual entity against all the other entities. We immediately appreciate the conceptual similarity between the present problem and the ranking tournament problem, which has been addressed by a variety of methods, e.g. Ali, Cook, and Kress (1986), Bordner (2016) and Luo, Xu, Zhang, and Zhang (2016), Cook and Hebner (1993), Goddard (1983). Here, we select a heuristic ranking procedure proposed by Cook, Golan, Kazakov, and Kress (1988) based on minimum violation ranking theory (Ali et al., 1986).[3] The modified heuristic procedure is described as follows.

Step 1. Rank the entities according to the strength score, i.e. row sums of the outranking matrix. If there are no ties, the procedure terminates. Otherwise, go to Step 2.

Step 2. Break the ties among any k entities by considering only their $k*k$ sub-matrix, calculating their strength scores, and then ranking these k entities based on their reduced scores.

Step 3. $k-1$ entities will be broken by performing Step 2.

Based on the above heuristic procedure, we can derive the same performance rankings of the entities in the numerical example as those given by Munda and Nardo (2009). Nevertheless, the computation procedure seems to be more efficient (see Table 2). While

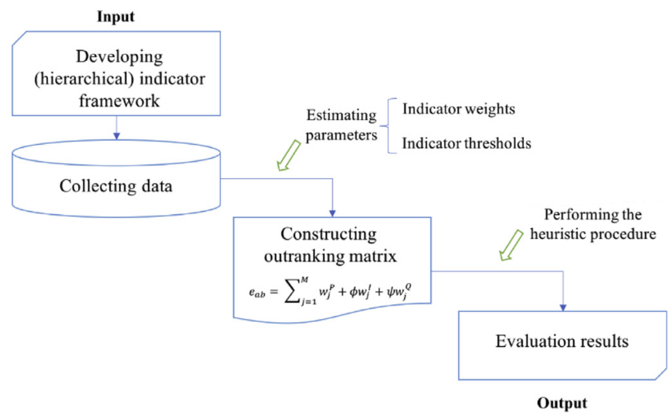

Fig. 2. Analytical procedure of the proposed non-compensatory approach.

the original non-compensatory framework is in need of calculating $n!$ possible rankings, we here only need to calculate n values if there is no tie among entities. If there exist k ($k < n$) ties after the first step of the heuristic procedure, we only need to calculate another k values. As a whole, we may get the overall performance ranking after calculating $n + k + \cdots$ values depending on the odds of the ties. An additional advantage of the heuristic procedure is that it can break the tie of rankings, which often appears when there are more entities.

In summary, the proposed non-compensatory approach for constructing CI can be graphically represented by Fig. 2.

3. Empirical study

3.1. Background

China had made remarkable achievements on the path towards a low-carbon society during the 12th Five-Year Plan (FYP) period (2011–2015). The ratios of non-fossil energy and natural gas consumption to total energy consumption increased by 2.6% and 1.9%, respectively. Meanwhile, the share of coal consumption dropped by 5.2%. The newly installed renewable energy generation capacity increased quickly and accounted for 40% of the world total. Moreover, the aggregate energy intensity and carbon intensity declined by 18.4% and 20%, respectively. China also completed afforestation of 2.96 million hectares and forest tending of 38.8 million hectares during the period. In her 13th FYP (2016-2020), China shows higher ambition on energy conservation and carbon reduction. According to the 13th FYP for Energy Development (NDRC, 2016b), China set a 15% minimum target for the share of non-fossil energy consumption in total energy consumption by 2020. Compared to the 2015 levels, China's energy intensity will be decreased by 15% and carbon intensity will drop by 18%.

Cities with the most intensive industrial activities and infrastructure construction undoubtedly play an important role for China to achive the goals and be transformed into a low-carbon society. Studies have evidenced that cities are responsible for a substantial share of carbon emissions in China (Dhakal, 2009). During

[3] A violdation occurs when entity A ranks over but performs worse than entity B. Minimum violation ranking theory is proposed to determine the minimum violation ranking when violation arises (Ali, Cook, & Kress, 1986).

Table 4
Indicators for low-carbon city performance evaluation.

Criteria	Indicator	Abbr.	Unit
Economic	[−] Energy intensity	EnergCon	Ton of SCE/10k yuan
	[+] R&D expenditure to GDP	R&D	%
	[+] Loan volume to GDP	Loan	%
	[+] Proportion of tertiary industry	TerInd	%
Living quality	[+] Proportion of public green space	PubSpa	%
	[−] Water consumption intensity	WatCon	L/captia/day
	[−] Engle's coefficient	Engle	%
	[−] Registered unemployment rate	Unemp	%
	[−] Population density	PopDen	Population/km^2
Environment	[+] Days of air quality equal to or above Grade II	GradeII	Days
	[+] Ratio of industrial solid wastes utilized	SolWas	%
	[+] Proportion of waste water treated	WasWater	%
	[+] Ratio of consumption waste treated	ConWas	%
Consumer behavior	[+] Public buses per capita	PubBus	Public buses/10k persons
	[+] Passengers intensity	PasInt	Times

the 13th FYP, about 100 million rural people will move to urban areas as permanent urban residents (NDRC, 2016a). The rural-to-urban migration improves China's urbanization level, while at the same time may bring cities more challenges, e.g. increasing energy demand, transportation congestion, urban sprawl and so on. In the circumstance, both central and local governments in China have recognized the need of more tangible and workable actions for tackling the challenge of carbon reductions. As a response, the National Development Reform Commission (NDRC) had initialized several pilot low-carbon programs. A total of six provinces and 36 cities were selected as the pioneers for low-carbon practices in 2010 and 2012. In the second session of China-U.S. Climate-Smart/Low-Carbon Cities Summit held in 2016, Chinese government announced that the number of pilot low-carbon cities are going to be increased to 100. Several cities (e.g. Beijing, Dalian, Shanghai, Nanjing, Wuhan, Chengdu, Guangzhou, Shenzhen and Hong Kong) have also taken part in the C40 networks which commit to address the problems caused by climate change.

Along the endeavor of low-carbon development, there is a need to perform a low-carbon performance assessment at city level in China. In the following sections, we shall apply the methodology presented in Section 2 to develop a non-compensatory CI for evaluating and comparing the low-carbon performance of forty cities in China.

3.2. Indicators

A set of 15 indicators has finally been selected for our empirical study, which are under four criteria such as economic, living quality, environment and consumer behavior. Table 4 shows the descriptions of the indicators. Generally, the indicator framework defined is similar to the one proposed by Tan et al. (2017). Additionally, in line with the suggestions by DTI (2003), we also attempt to incorporate other key aspects, e.g. quality of life, technological innovation, job opportunities, into the framework. Compared to Tan et al. (2017), several indicators have been modified to cater for the nature of the status of Chinese cities. For instance, three indicators, i.e. the ratios of energy consumption to GDP, R&D expenditure to GDP and loan volume to GDP, are used to replace per capita GDP to evaluate cities economic performance in a more comprehensive way. The main reason is that the three indicators have frequently appeared in Chinese government reports. Furthermore, the ratio of energy consumption to GDP, as a monetary-based indicator, is more suitable for evaluating energy efficiency at a high level of aggregation (Ang, 2006). The R&D expenditure could be treated as a proxy for technological innovation. Due to the lack of data, the indicator "days of air quality equal to or above Grade II" in environment dimension is taken as the substitute for the indicators measuring the levels of main air pollutants such as NO, SO, PM2.5 and PM10. In order to measure the quality of life and job opportunities, we incorporate Engle's coefficient and unemployment rate into living quality criteria by following the suggestions by Yu (2014). Moreover, passenger intensity, defined as the ratio of the indicator "passengers transported by buses trolley buses" to the indicator "annual average population", is included to represent consumer low-carbon transport behavior.

3.3. Data and results

We apply the non-compensatory approach to evaluating the low-carbon performance of forty Chinese cities based on the indicator framework described in Section 3.2. Whether a city is listed in the national low carbon program is the main principle for selecting sample cities. It should be pointed out that these selected forty cities are not exactly the same as the aforementioned pilot cities, because of the lack of data for several pilot cities such as Hulun Buir, Chizhou and Jingdezhen. In order to achieve a comprehensive comparison as far as possible, several provincial capital cities, e.g. Hefei, Jinan, Fuzhou etc., have also been included.

The data on the indicators for the year 2014 are collected from several sources including *China City Statistical Yearbook*, *China Statistical Yearbook on Environment*, and *Municipal and Regional Statistical Yearbook*. The data is generally for urban areas. When the data in statistical yearbook does not differentiated between urban and rural areas, the overall level is taken as the proxy for the corresponding indicator.

In order to apply the approach described in Section 2, we need to determine the weights and threshold values for the indicators. Weights of indicators are suggested to have a significant effect on the overall results in term of composite indicator. JRC (2008) summarized several weighting methods, e.g. data-driven (principal components analysis and data envelopment analysis) and normative (equal weighting, analytic hierarchy process and budget allocation) methods. Each weighting method has its own merits. However, a possible misunderstanding is that the weight assigned to an indicator can be directly interpreted as the importance coefficient of the indicator, except for the weights derived from budget allocation and equal weighting (Becker, Saisana, Paruolo, & Vandecasteele, 2017). When there is a lack of decision makers or expert panel, equal weighting seems to be a fairly reasonable choice for determining indicator weights in order to keep their consistency with the meaning of importance coefficient. Other advantages of equal weighting method are its transparency and fairly good compatibility with all kinds of aggregation methods (JRC, 2008). As such, in our empirical study we take the equal weighting method for use. Considering that there are two levels in

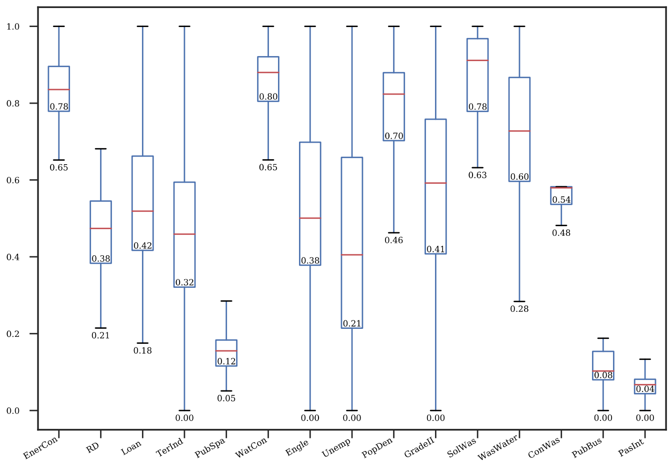

Fig. 3. Comparative box plots of the indicator performance based on normalized data.

the indicator framework, we hierarchically assign indicator weight, i.e. we firstly assign each criterion $\frac{1}{\mathcal{N}}$ as its weight where $\mathcal{N}$ is the number of criteria (=4). Then, the indicator weight is determined by the ratio of the criterion weight to the number of indicators under the criterion. For example, each indicator under economic criterion would be assign the value $\frac{1}{4} * \frac{1}{4} = \frac{1}{16}$ as its weight.

With respect to indicator thresholds, analysts are usually suggested to collaborate closely with decision-makers in the process of expressing preference judgments (Attardi et al., 2018; Bottero, Ferretti, Figueira, Greco, & Roy, 2015; Roy et al., 2014). It requires analysts to explain the nature of the data and provide an clarified instruction so that decision-makers can follow the instruction and use their own experience to provide a reliable sample of thresholds for the indicators. When decision-makers are unfamiliar with the theory of the evaluation method, analysts may provide a threshold reference set for both indifference and preference thresholds. For example, Attardi et al. (2018) arbitrarily defined 10% and 20% of the maximum difference between entities with respect to indicators as the reference set for indifference and preference thresholds, respectively. Because the determination of thresholds is heavily dependent on the nature of each indicator (Roy et al., 2014), we here intend to provide an alternative way for deriving indicator thresholds. We firstly examine the statistical distribution of the indicator values by a box-and-whisker plot shown in Fig. 3. Due to the characteristics of box-and-whisker plot, the normal data should be included between the whiskers. The values lying out of the range are the outliers. If we define the Lower Extreme and the Lower Quartile as the indifference and preference threshold of each indicator, we can partially exclude the impact of outliers on the overall performance evaluation. Base on the reasoning, we choose the Lower Extreme and the Lower Quartile as the indifference and preference threshold for each indicator.

In addition to weights and thresholds of each indicator, the coefficient of weak preference relation may also have an important impact on the overall low-carbon performance evaluation due to its significant role in constructing outranking matrix. Because of the characteristics of weak preference, we are in the presence of an ambiguity situation relevant to a hesitation between two conclusions, i.e. indifference and preference. Hence, we split the coefficient of weak preference relation into three scenarios by setting $\psi = \frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{3}$, respectively, and investigate its influence on the overall low-carbon performance evaluation.

Using the proposed non-compensatory CI approach and the parameter settings described above, we obtain the results of low-carbon performance evaluation for forty cities in China. Fig. 4 shows the low-carbon performance ranking map of the cities with different thresholds. For comparison purpose, we also provide the ranking map without consideration of indicator thresholds in Fig. 4.

It can be seen from Fig. 4 that the coastal cities in China generally performed better than the other regions on low carbon

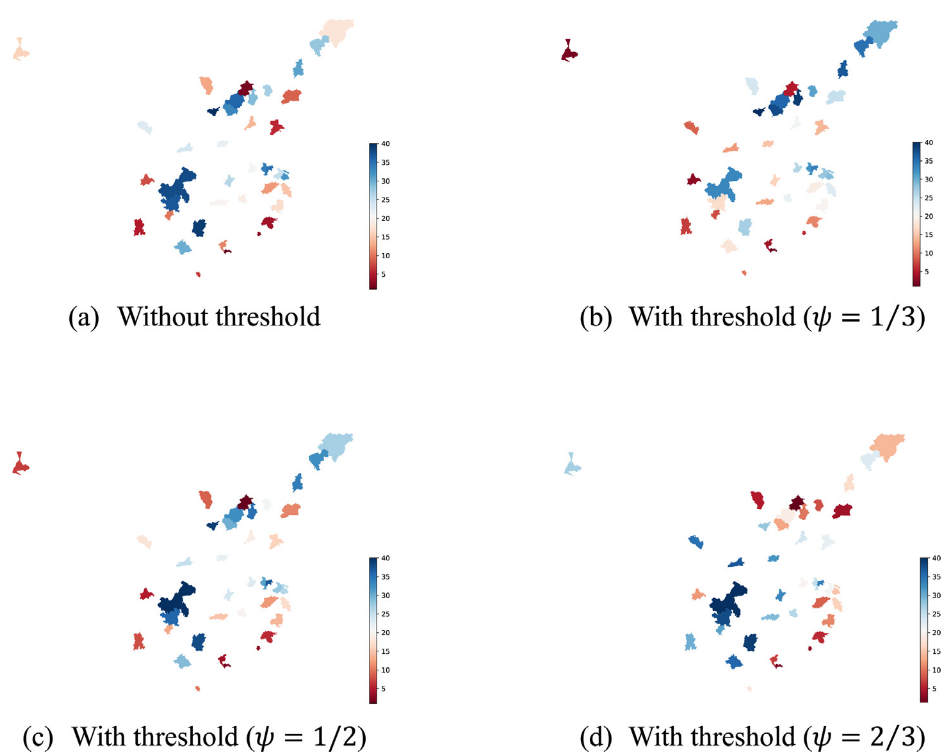

Fig. 4. Low-carbon performance ranking map of forty cities in China.

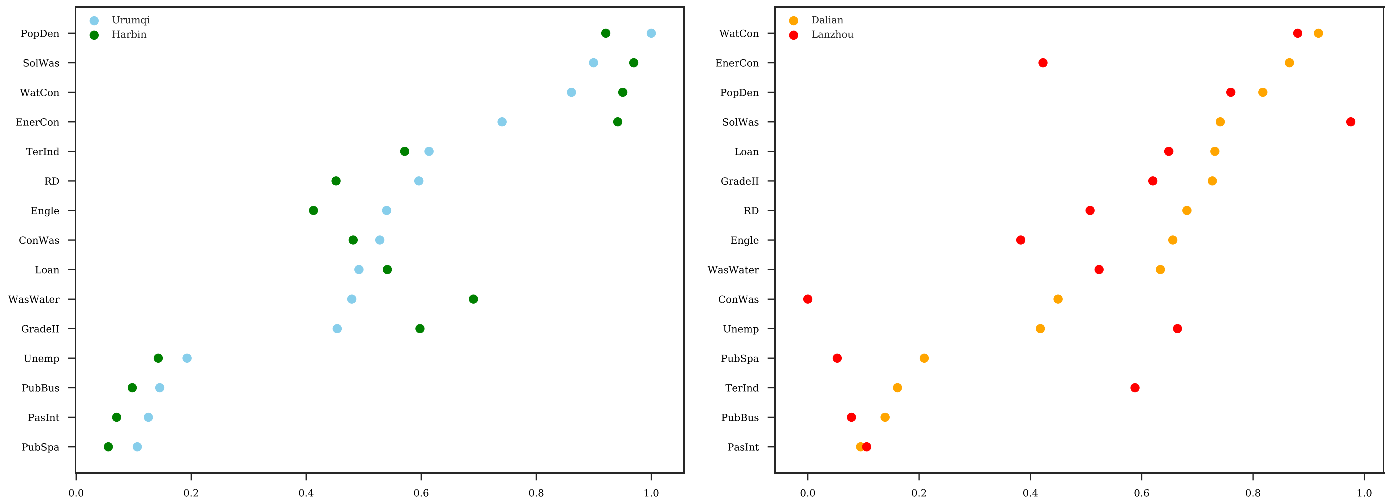

Fig. 5. Indicator performance difference of cities based on normalized data.

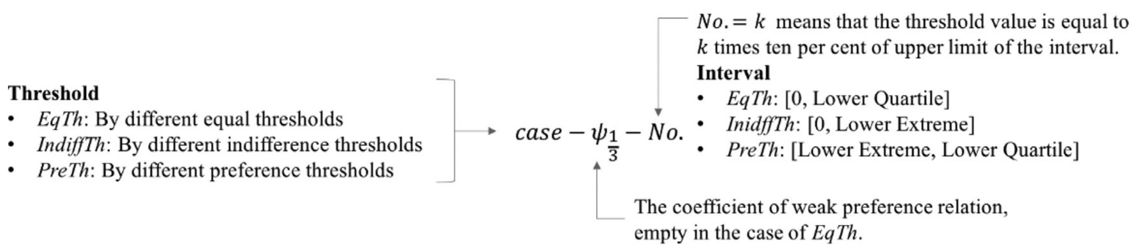

Fig. 6. Cases for sensitivity analysis.

development, no matter what parameter combination is used. Several western cities, such as Urumqi, Chengdu, Kunming and Zunyi, also have good low-carbon performance. On the contrary, the mid-western cities and the cities nearby Beijing generally have poor low-carbon performance. It may be an indication that decision-makers in those cities, especially for the pilot low-carbon cities such as Chongqing, Baoding, Shijiazhuang, Tianjin and Wuhan, should pay more attention to low-carbon city development.

Compare Fig. 4(a) with other three sub-figures in Fig. 4, we may conclude that the indicator thresholds have a significant impact on the overall low-carbon performance evaluation results. Nevertheless, we cannot directly judge whether the overall low-carbon performance evaluation with consideration of indicator thresholds is more consistent with the reality. We further evaluate the empirical results with and without thresholds by referring to earlier relevant studies such as Liu (2016). Considering the actual emissions of different regions in China as reported by Liu (2016), we may conclude that it is necessary to take indicator thresholds into account to provide more consistent evaluation results with actual situations in applying the proposed non-compensatory approach to evaluate low-carbon performance.

To investigate the impact of the coefficient of weak preference on overall performance evaluation results, we turn back to the indicator values by using several selected cities. Firstly, we analyze the indicator performance difference between Urumqi and Harbin (see the left sub-figure in Fig. 5). It is found that Harbin is dominated by Urumqi in nine indicators, and at the same time Urumqi is dominated by Harbin in the remaining six indicators. Taking indicator thresholds into account, we find that the low-carbon performance of Urumqi is at least as good as that of Harbin in all indicators. Hence, we believe that ranking Urumqi over Harbin on low-carbon performance is more consistent with the indicator values. By the reasoning, we suggest that the coefficient of the weak preference relation should not be greater or equal to 2/3. Through pair-wise comparison of other cities, e.g. Shenyang, Jilin, Lanzhou, Wuhan and Xi'an, we can reach a similar conclusion. Fig. 5 also shows the indicator performance difference between Dalian and Lanzhou. Clearly, ranking Dalian over Lanzhou on low-carbon performance seems to be more consistent with the indicator values. This conclusion may indirectly imply that the coefficient of the weak preference relation should not be less or equal to 1/3. In summary, it is necessary to consider indicator thresholds in evaluating low-carbon city performance by the proposed non-compensatory CI approach. In the current empirical study, we find that more consistent evaluation results can be obtained by setting the coefficient of the weak preference relation as 1/2.

3.4. Sensitivity analysis

Previous analysis is based on that the indifference and preference thresholds of each indicator are fixed at the Lower Extreme and Lower Quartile of its statistical distribution. Because of the absence of decision-makers or expert panel, we intend to investigate the influence of different indicator thresholds on the performance evaluation results by considering three main threshold cases, i.e. by considering different thresholds when indifference threshold is equal to corresponding preference threshold (hereafter denoted by *EqTh*), by considering different indifference thresholds while keeping corresponding preference threshold unchanged (*IndiffTh*), and by considering different preference thresholds while keeping corresponding indifference thresholds unchanged (*PreTh*). Every indifferent and preference threshold will gradually be increased by a certain rate in the specific interval. For example, when discussing the *EqTh* case, we assume that its lowest and highest values are respectively zero and Lower Quartile. For the purpose of clarification, we denote the cases by a symbol like $case - \psi_{1/3} - No.$ which is explained in detail in Fig. 6.

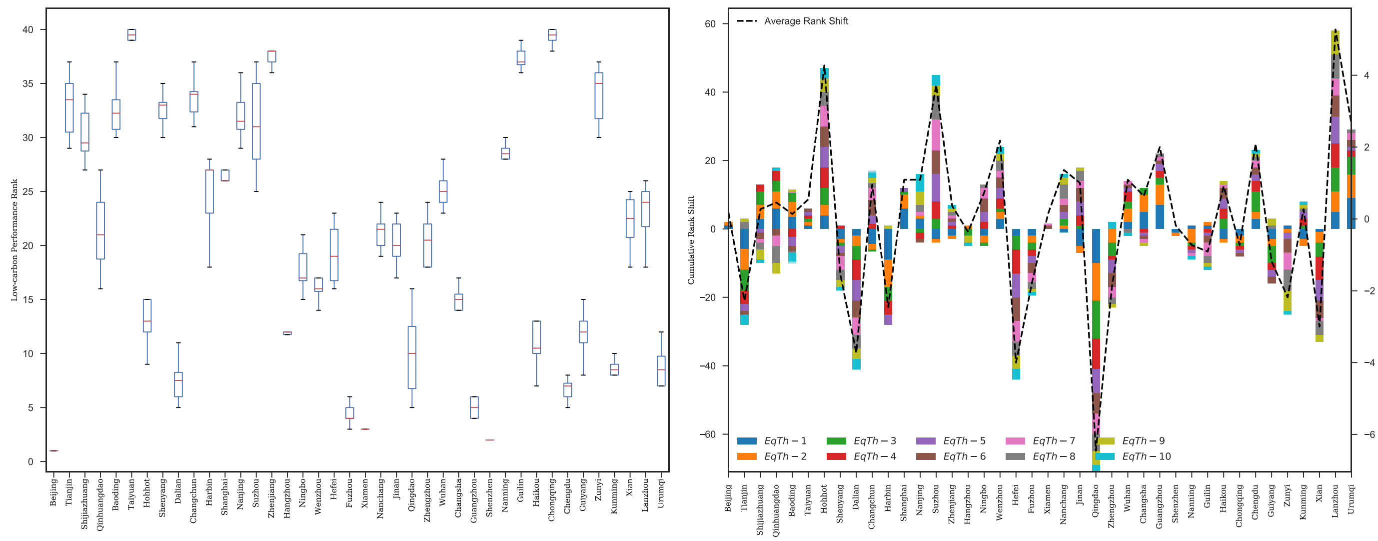

Fig. 7. The rank variance and rank shift with different equal thresholds.

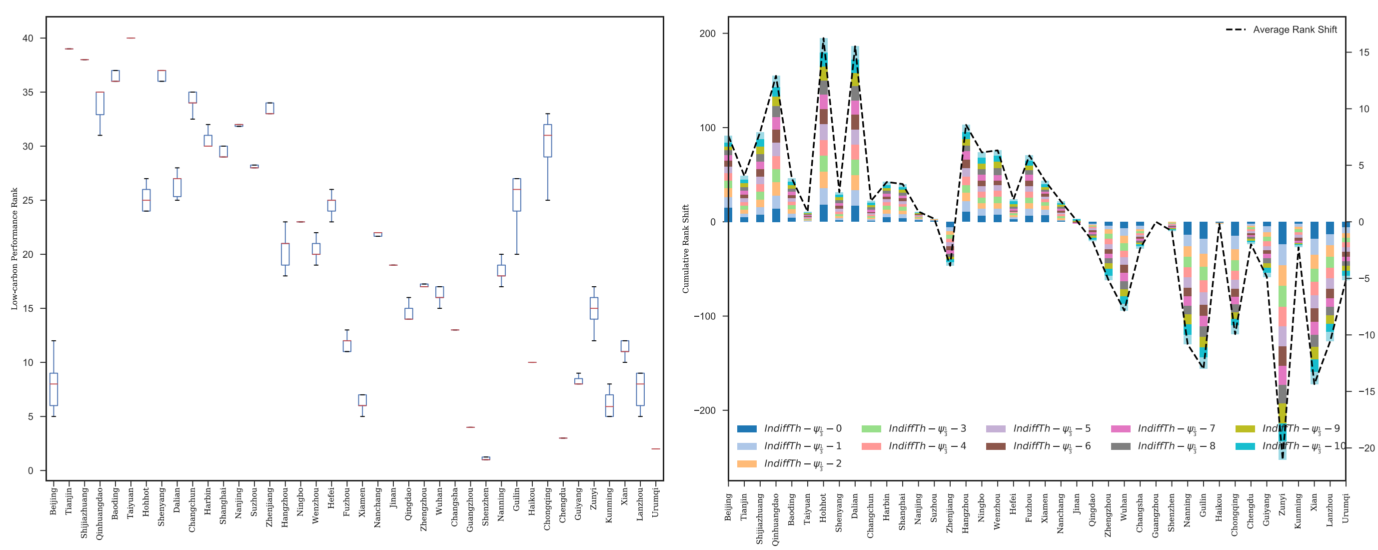

Fig. 8. The rank variance and rank shift with different indifference thresholds.

Based on the previous analysis, we here only focus on two scenarios of the coefficient of the weak preference relation, i.e. $\psi = 1/3$ and $\psi = 1/2$. Since the coefficient of weak preference relaxes the impact of the indifference threshold on the outranking matrix when $\psi = 1/2$, we will incorporate this situation into the *EqTh* case. For every city, we can get a number of ranks for every case. The comparable box plots of their ranks are provided in the left parts of Figs. 7 and 8. We also explore the cumulative shift and the average shift in city rankings which are showed in the right parts of Figs. 7 and 8. The cumulative rankings shift (denoted by R_c) is calculated by $R_c = \sum_{n=1}^{N}(Rank_{ref} - Rank)$, where N is the total number of cities, $Rank_{ref}$ denotes the reference ranking given in Section 3.3, and *Rank* is the ranking obtained for each case. The average ranking shift is equal to the ratio of R_c to the total number of the rankings for each case. Note that the way for calculating cumulative ranking shift here is similar to that used by JRC (2008), although the absolute differences are used in the latter. The main reason is that we want to explore the ranking shift direction of each city with respect to the change of threshold.

Comparing the rankings under different cases, we find that in the *IndiffTh* case the CI shows more robust behavior. However, in the *PreTh* case several cities' rankings change dramatically, e.g. from 8 to 24 for Hohhot, from 7 to 25 for Dalian, and from 16 to 36 for Zunyi. The dramatic change of rankings in the *PreTh* case implies that the preference threshold have a significant impact on the low-carbon performance evaluation results. The right part of Fig. 7 also supports the conclusion. Nevertheless, based on the three comparative box plots, we find that the rankings of leaders such as Beijing, Hangzhou, Shenzhen, Chengdu, and Urumqi, and laggards such as Taiyuan, Zhenjiang and Chongqing are only slightly influenced by the indicator thresholds in all three cases.

The right parts of Figs. 7 to 9 show that almost all cities' ranking shifts along one direction with respect to different indicator thresholds in all three cases, except for Baoding, Dalian, Hefei, Fuzhou, Lanzhou and Urumqi in the case of *PreTh*. The rankings of these cities generally decrease first along the increase of preference threshold and increase later. A possible explanation is that the preference relation of these cities is inversed to the weak preference with respect to some indicators when the preference thresholds increase to certain values. We also find that both the cumulative rank shift and the average shift in the *indiffTh* case change dramatically, while show relatively modest changes in other two

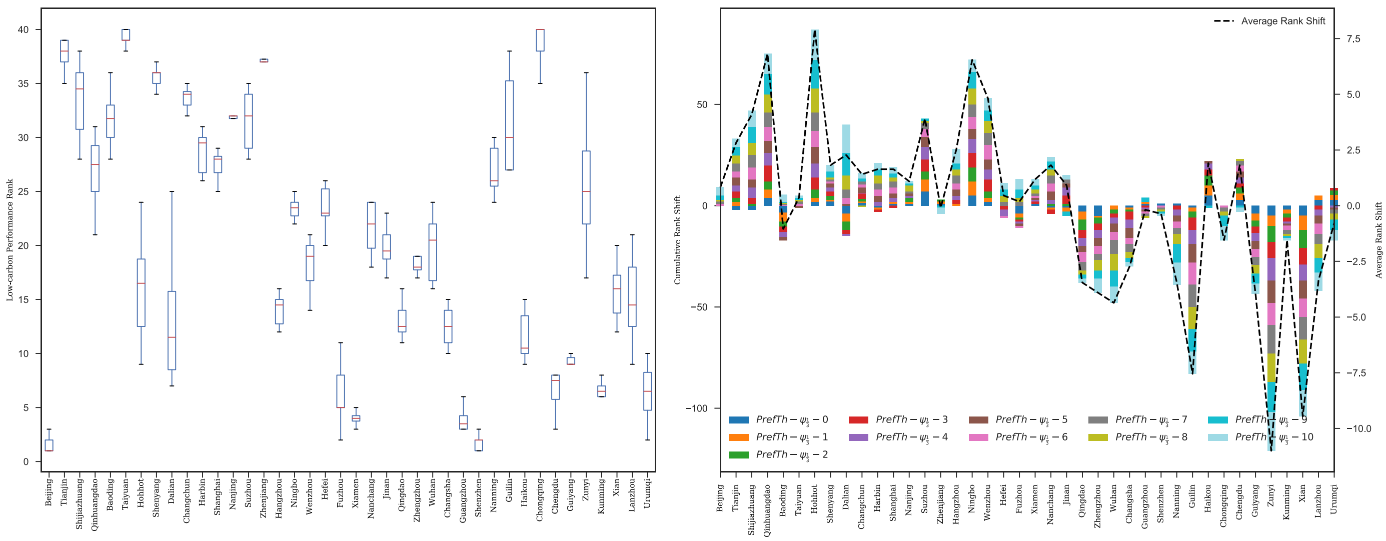

Fig. 9. The rank variance and rank shift with different preference thresholds.

cases. In summary, different indicator thresholds will generate different evaluation results. If the indicator thresholds are assigned by decision-maker or expert panel, the assignment procedure should be elaborated clearly and transparently to increase the robustness and usefulness of the proposed non-compensatory CI approach.

4. Conclusion

This paper contributes to develop a non-compensatory approach to constructing CI for evaluating low-carbon performance. A CI is said to be non-compensatory if such model does not allow the weakness of one indicator to be compensated by another indicator's strengths. The proposed approach has the flexibility of integrating mixed quantitative and qualitative indicators and allows for these measures to be reported in their natural units without any transformation, which seems to be more informative and intuitive to policy makers and public. Compared to the non-compensatory multi-criteria framework proposed by Munda and Nardo (2009), the approach presented in this paper is computationally more efficient at the stage of data aggregation. Through integrating thresholds, the approach can avoid the Luce's paradox of "the cup of coffee" in performance comparison.

The proposed approach has been applied to evaluate low-carbon performance for forty Chinese cities. The results show that the coastal cities in China generally performed better than the cities in other regions on low carbon development. Some western cities, such as Urumqi, Chengdu, Kunming and Zunyi, also have good low-carbon performance. On the contrary, the midwestern cities and the cities nearby Beijing showed poor low-carbon performance. In the empirical study, we have attempted to assign the reference set of indifference and preference thresholds by reference to the statistical distribution of the indicator values in order to avoid the arbitrariness in their determinination. Our empirical study also shows that the proposed non-compensatory approach with the coefficient of the weak preference relation equal to ½ may generate more reasonable evaluation results. We have also investigated the impact of different thresholds on the cities' low-carbon performance rankings. It is found that the overall low-carbon performance ranking of a certain city is quite sensitive to the parameters, especially to the preference threshold. We thus suggest that the procedure for determining thresholds should be elaborated clearly and transparently to increase the robustness of the proposed non-compensatory CI, especially when the thresholds are assigned by decision-maker or expert panel.

Despite the usefulness of the proposed non-compensatory CI approach in evaluating the low carbon performance of Chinese cities, it has inevitably some limitations. First, for the purpose of transparency, comparability over time and space as well as ease of interpretation, we apply equal weighting method to assign the indicator weights. It is worthwhile developing an alternative weighting scheme that can take into account the regional difference and further improve the robustness of the proposed approach. Second, although mutiple scenarios are considered in the empirical study, it may still be subjective in determining the coefficient of the weak preference relation for constructing outranking matrix. To the best of our knowledge, it is very difficult to assess whether a coefficient of the weak preference is good enough. The method used in this paper, i.e. trace back to the indicator data and make a comparison with relevant studies and realistic situations, may be a feasible way to justify the reasonableness of a certain value. Nevertheless, we suggest that more scrutiny on this point should be taken in future research. Third, preliminary statistical analysis may be performed to identify data outliers. Once the outliers are identified, extra efforts should be made to better clarify the robustness of the proposed approach with respect to outliers, e.g. incorporating the weighting scheme proposed by Vidoli, Fusco, and Mazziotta (2015). Finally, it should be pointed out that our empirical study attempts to uncover the status of low-carbon development in Chinese cities from an individual perspective. Further research may be carried out by considering the spatial autocorrelation among cities in order to draw a more complete picture of low-carbon development in China.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Nos. 71625005 and 715731197) and the Funding of Jiangsu Innovation Program for Graduate Education (KYZZ16_0159).

References

Ali, I., Cook, W. D., & Kress, M. (1986). On the minimum violations ranking of a tournament. *Management Science, 32*, 660–672.
Ang, B. W. (2006). Monitoring changes in economy-wide energy efficiency: From energy – GDP ratio to composite efficiency index. *Energy Policy, 34*, 574–582.
Attardi, R., Cerreta, M., Sannicandro, V., & Torre, C. M. (2018). Non-compensatory composite indicators for the evaluation of urban planning policy: The Land-Use

Policy Efficiency Index (LUPEI). *European Journal of Operational Research, 264*, 491–507.
Becker, W., Saisana, M., Paruolo, P., & Vandecasteele, I. (2017). Weights and importance in composite indicators: Closing the gap. *Ecological Indicators, 80*, 12–22.
Blancas, F. J., Caballero, R., González, M., Lozano-Oyola, M., & Pérez, F. (2010). Goal programming synthetic indicators: An application for sustainable tourism in Andalusian coastal counties. *Ecological Economics, 69*, 2158–2172.
Bottero, M., Ferretti, V., Figueira, J. R., Greco, S., & Roy, B. (2015). Dealing with a multiple criteria environmental problem with interaction effects between criteria through an extension of the Electre III method. *European Journal of Operational Research, 245*, 837–850.
Bordner, S. S. (2016). 'All-things-considered', 'better-than' and sports rankings. *Journal of the Philosophy of Sport, 43*, 215–232.
Böhringer, C., & Jochem, P. E. P. (2007). Measuring the immeasurable – A survey of sustainability indices. *Ecological Economics, 63*, 1–8.
Cook, W. D., Golan, I., Kazakov, A., & Kress, M. (1988). A case study of a non-compensatory approach to ranking transportation projects. *Journal of the Operational Research Society, 39*, 901–910.
Cook, W. D., & Hebner, K. J. (1993). A multicriteria approach to country risk evaluation: With an example employing Japanese data. *International Review of Economics & Finance, 2*, 327–348.
Despotis, D. K. (2005). Measuring human development via data envelopment analysis: The case of Asia and the Pacific. *Omega, 33*(5), 385–390.
Dhakal, S. (2009). Urban energy use and carbon emissions from cities in China and policy implications. *Energy Policy, 37*, 4208–4219.
Díaz-Balteiro, L., & Romero, C. (2004). In search of a natural systems sustainability index. *Ecological Economics, 49*, 401–405.
Diaz-Balteiro, L., González-Pachón, J., & Romero, C. (2017). Measuring systems sustainability with multi-criteria methods: A critical review. *European Journal of Operational Research, 258*(2), 607–616.
DTI (Department of Trade and Industry). (2003). *UK energy white paper: Our energy future – Creating a low carbon economy*. London: TSO.
Ebert, U., & Welsch, H. (2004). Meaningful environmental indices: A social choice approach. *Journal of Environmental Economics and Management, 47*, 270–283.
Fusco, E. (2015). Enhancing non-compensatory composite indicators: A directional proposal. *European Journal of Operational Research, 242*, 620–630.
Fusco, E., Vidoli, F., & Sahoo, B. K. (2017). *Spatial heterogeneity in composite indicator: A methodological proposal*. Omega in press.
Goddard, S. T. (1983). Ranking in tournaments and group decision making. *Management Science, 29*, 1384–1392.
Govindan, K., & Jepsen, M. B. (2016). ELECTRE: A comprehensive literature review on methodologies and applications. *European Journal of Operational Research, 250*, 1–29.
IPCC (Intergovernmental Panel on Climate Change). (2015). *Climate change 2014: Mitigation of climate change*. Cambridge University Press.
JRC (Joint Research Centre - European Commission). (2008). *Handbook on constructing composite indicators: Methodology and user guide*. OECD Publishing.
Liu, Z. (2016). *China's carbon emissions report 2016: Regional carbon emissions and the implication for China's low carbon development*. Belfer Center for Science and International Affairs Technical Report.
Luce, R. D. (1956). Semiorders and a theory of utility discrimination. *Econometrica, 24*, 178–191.
Luo, K., Xu, Y., Zhang, B., & Zhang, H. (2016). Creating an acceptable consensus ranking for group decision making. *Journal of Combinatorial Optimization* https://doi.org/10.1007/s10878-016-0086-9.
Mori, K., & Christodoulou, A. (2012). Review of sustainability indices and indicators: Towards a new city sustainability index (CSI). *Environmental Impact Assessment Review, 32*, 94–106.
Munda, G., & Nardo, M. (2005). *Constructing consistent composite indicators: The issue of weights* Technical Report, Institute for the Protection and Security of the Citizen, European Commission. Available at: *http://publications.jrc.ec.europa.eu/repository/handle/JRC32434*.
Munda, G., & Nardo, M. (2009). Noncompensatory/nonlinear composite indicators for ranking countries: A defensible setting. *Applied Economics, 14*, 1513–1523.
NDRC (National Development and Reform Commission). (2016a). *The 13th five-year plan*. National Development and Reform Commission *http://en.ndrc.gov.cn/newsrelease/201612/P020161207645765233498.pdf*.
NDRC (National Development and Reform Commission). (2016b). *The 13th five-years plan for energy development*. National Development and Reform Commission *http://www.ndrc.gov.cn/zcfb/zcfbghwb/201701/W020170117350627940556.pdf*.
Pollesch, N., & Dale, V. H. (2015). Applications of aggregation theory to sustainability assessment. *Ecological Economics, 114*, 117–127.
Price, L., Zhou, N., Fridley, D., Ohshita, S., Lu, H., Zheng, N., & Fino-Chen, C. (2013). Development of a low-carbon indicator system for China. *Habitat International, 37*, 4–21.
Rogge, N. (2018). On aggregating benefit of the doubt composite indicators. *European Journal of Operational Research, 264*, 364–369.
Rowley, H. V., Peters, G. M., Lundie, S., & Moore, S. J. (2012). Aggregating sustainability indicators: Beyond the weighted sum. *Journal of Environmental Management, 111*, 24–33.
Roy, B. (1991). The outranking approach and the foundations of ELECTRE methods. *Theory and Decision, 31*, 49–73.
Roy, B., Figueira, J. R., & Almeida-Dias, J. (2014). Discriminating thresholds as a tool to cope with imperfect knowledge in multiple criteria decision aiding: Theoretical results and practical issues. *Omega, 43*, 9–20.
Tofallis, C. (2013). An automatic-democratic approach to weight setting for the new human development index. *Journal of Population Economics, 26*, 1325–1345.
Tofallis, C. (2014). On constructing a composite indicator with multiplicative aggregation and the avoidance of zero weights in DEA. *Journal of the Operational Research Society, 65*, 791–792.
Tan, S., Yang, J., Yan, J., Lee, C., Hashim, H., & Chen, B. (2017). A holistic low carbon city indicator framework for sustainable development. *Applied Energy, 185*, 1919–1930 Part 2..
Truong, T. D., Adamowicz, W. L. V., & Boxall, P. C. (2015). Modeling non-compensatory preferences in environmental valuation. *Resource and Energy Economics, 39*, 89–107.
Yoon, K. P., & Hwang, C. L. (1995). *Multiple attribute decision making: An introduction*: 104. Sage Publications.
Yu, L. (2014). Low carbon eco-city: New approach for Chinese urbanization. *Habitat International, 44*, 102–110.
Van Puyenbroeck, T., & Rogge, N. (2017). Geometric mean quantity index numbers with benefit-of-the-doubt weights. *European Journal of Operational Research, 256*, 1004–1014.
Verbunt, P., & Rogge, N. (2018). Geometric composite indicators with compromise benefit-of-the-doubt weights. *European Journal of Operational Research, 264*, 388–401.
Vidoli, F., Fusco, E., & Mazziotta, C. (2015). Non-compensability in composite indicators: A robust directional frontier method. *Social Indicators Research, 122*, 635–652.
Zanella, A., Camanho, A. S., & Dias, T. G. (2015). Undesirable outputs and weighting schemes in composite indicators based on data envelopment analysis. *European Journal of Operational Research, 245*, 517–530.
Zhou, N., He, G., Williams, C., & Fridley, D. (2015). ELITE cities: A low-carbon eco-city evaluation tool for China. *Ecological Indicators, 48*, 448–456.
Zhou, P., Ang, B. W., & Poh, K. L. (2006). Comparing aggregating methods for constructing the composite environmental index: An objective measure. *Ecological Economics, 59*, 305–311.
Zhou, P., Ang, B. W., & Poh, K. L. (2007). A mathematical programming approach to constructing composite indicators. *Ecological Economics, 62*, 291–297.
Zhou, P., Ang, B. W., & Zhou, D. Q. (2010). Weighting and aggregation in composite indicator construction: A multiplicative optimization approach. *Social Indicators Research, 96*, 169–181.
Zhou, P., Ang, B. W., & Wang, H. (2012). Energy and CO_2 emission performance in electricity generation: A non-radial directional distance function approach. *European Journal of Operational Research, 221*, 625–635.
Zhou, P., Delmas, M., & Kohli, A. (2017). Constructing meaningful environmental indices: A nonparametric frontier approach. *Journal of Environmental Economics and Management, 85*, 21–34.